2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

HAGrasp: Hybrid Action Grasp Control in Cluttered Scenes using Deep Reinforcement Learning

Kai-Tai Song and Hsiang-Hsi Chen, *Member, IEEE*

***Abstract*—Robotic autonomous grasp requires the system to perform multiple functions such as gripper and robot control, making it a task with hybrid output nature. Existing methods based on closed-loop deep reinforcement learning rely on external models for termination evaluation. To achieve more effective grasp for novel objects, we propose a new autonomous grasp control scheme termed HAGrasp that considers the complete point cloud of the workspace. It integrates grasp pose estimation, end-effector pose evaluation, and motion planning of the robotic arm into a single model, enhancing the success rate while reducing computational load. We present a closed-loop grasp control system based on deep reinforcement learning. This control system can perform grasp tasks while dynamically adjusting to avoid end-effector collisions. The design of hybrid-action reinforcement learning module is trained with unified latent action space and further improve generalization, achieving real-time autonomous grasp control. Real robot experiments show that our method has 74.2% success rate for grasping 7 unseen objects. Comparative experiments show that the proposed HAGrasp outperforms open-loop baseline Contact-Graspnet in both success rate and inference time. It is demonstrated that with integrated multi-view input and sim-to-real training design, our method improves real-world applications of autonomous grasp.**

INTRODUCTION

Robotic autonomous grasp requires the system to perform gripper and robot control, making it a task with hybrid output nature [1]-[3]. For grasp tasks in cluttered environments, the robot needs to simultaneously perform obstacle avoidance and grasp estimation. Open-loop approaches [4]-[8], primarily focus on grasp estimation, treating motion planning as a minor problem. Most existing closed-loop methods are based on deep reinforcement learning (DRL) to handle low-level control but rely on separate models for termination evaluation [9], [10].

Grasp determination plays an important role in robotic grasp, where an efficient control system can notably mitigate computational load and cycle time. The utilization of open-loop methodologies combined with deep learning is constrained by performance bottleneck arising from single-view inputs [11]. This limitation becomes evident when objects are occluded due to their arbitrary orientations, leading to a decline in system performance. Furthermore, the absence of an additional motion planner can give rise to grasp failures caused by collisions between the robot and its surroundings.

* Research supported by National Science and Technology Council of Taiwan R.O.C. under grant MOST 112-2221-E-A49-112-.

Kai-Tai Song* is with the Institute of Electrical and Control Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan. (Corresponding author: 886-3-5731865, e-mail: ktsong@nycu.edu.tw)

Hsiang-Hsi Chen is with the Graduate Degree Program of Robotics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan (e-mail: a310605024.en10@nycu.edu.tw)

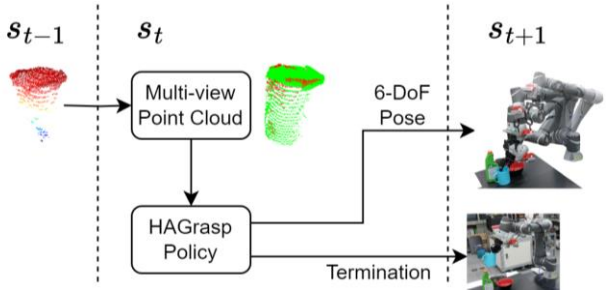


Fig. 1. Illustration of HAGrasp method. The policy takes multi-view inputs, and outputs hybrid action for gripper and end-effector control.

Typical closed-loop designs rely on policy learning to directly control robot actions as model outputs, but the termination condition is determined by an external model [9], [10], [12]. This decoupling of the termination evaluation from the training process leads to rewards obtained from interaction with the environment that are not directly related to the grasp task. As a result, the model struggles to understand the multimodal nature of grasp poses effectively. To enhance system performance and efficiency, it is a better practice to combine termination evaluation and end-effector control using a hybrid action control system.

In this work, we present Hybrid Action Grasp control system (HAGrasp), a system that performs collision avoidance and grasp simultaneously by leveraging the geometric features of point clouds. The function of robotic autonomous grasp is divided into two parts: waypoint prediction and termination evaluation. To integrate multiple outputs into a single model, termination evaluation is treated as a learnable parameter. As shown in Fig. 1, in the hybrid-output control approach, end-effector control is a continuous action in the 6-DoF Cartesian coordinate space, while termination is a binary discrete action that control action of the parallel gripper.

Further, robotic task with high-dimensionality and sparse reward raises the difficulty for RL training. Moreover, hybrid action in DRL design without further optimization enlarges performance drop in real world due to sim-to-real gap [2]. To cope with this problem, latent action space has been brought out to enable faster learning with an efficient action representation [10], [13], [14]. Our approach addresses this problem of the action space by using latent action from Conditional Variational Autoencoder (CVAE) [15]. The proposed method incorporates both outputs of discrete-continuous action space [16] into the optimization process of policy learning. An encoder is used to map continuous and discrete actions into a unified representation space for optimization. By incorporating termination evaluation into the Q-learning design, the model is optimized for end-effector control while considering the evaluation and quality of target grasp poses. This approach allows the model to learn the multi-modality and features of successful grasp.

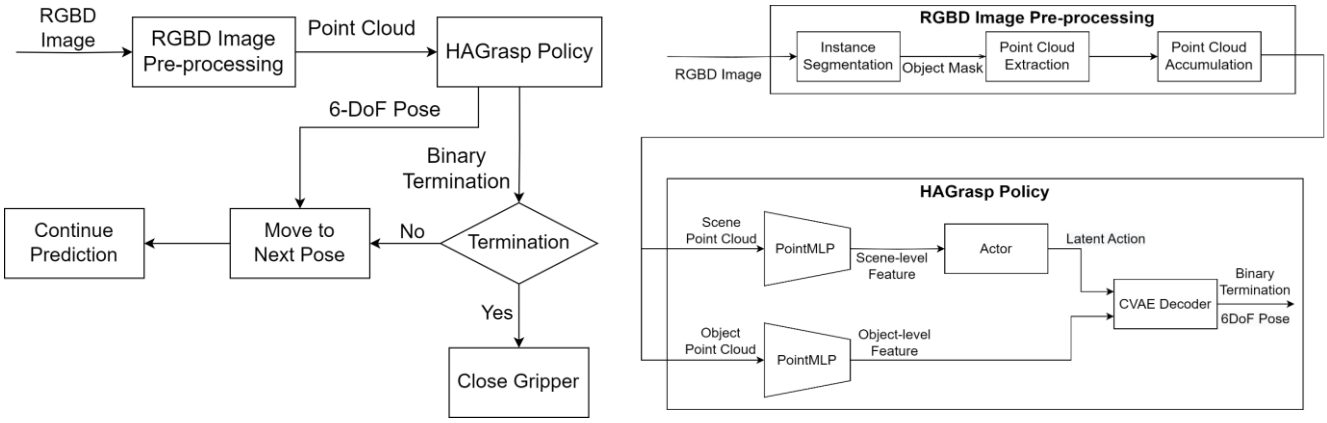


Fig. 2. System architecture of the proposed HAGrasp for novel-object grasping in clutter.

As shown in Fig. 2, the HAGrasp takes both scene and object features as input, making the policy has the capability of scene-level collision awareness while reconstructing the original action space based on the object-level point cloud. The system dynamically adjusts the end-effector and gripper pose while simultaneously evaluating whether the current pose satisfies the termination condition. At each waypoint, the system accumulates environment information through multi-view point cloud fusion, maximizing the performance of the control module. The continuous refinement of the end-effector pose and the evaluation of the termination condition contribute to maximizing the efficiency and success rate of the control module during the grasp task. The contributions of this paper can be summarized as follows:

- A hybrid action grasping framework that executes end-effector control and termination evaluation without colliding with objects in a complex scene.
- We propose a new action-mapping conditional variational encoder framework that projects hybrid continuous-discrete robot actions to latent ones with unified representation for hybrid action policy training.
- A novel data collection strategy that jointly training with reinforcement learning policy for fine-tuning action-mapping CVAE.
- A domain adaptation scheme is developed for point-cloud-based estimation in grasp training design. The performance drop in real-robot experiments is reduced and the performance that is close to simulated results is achieved.

I. Related Work

A. Open-Loop Collision-Free Grasp

In the design of open-loop grasp, single-view point clouds are used for grasp pose estimation. Some studies [5] have incorporated optimization mechanisms for grasp poses. To strengthen the model's attention on geometric characteristics during estimation, works [4], [6] simplify the form of grasp pose relationship between the gripper and the object, enabling the convergence of sparsely and imbalanced training objectives in the six-dimensional space towards reasonable object surfaces. Two-stage estimation method has been proposed to improve the performance [7], it predicts the grasp pose for the entire scene based on distribution of graspable regions. To address the shortcomings in handling small objects,[8] introduces a scale balanced loss function and multi-scale grasp pose search.

B. Closed-Loop Robot Grasp

RL-based approaches show promising results in achieving more robust and versatile capabilities for robots. End-to-end policy learning methods leverage extensive data to learn closed-loop grasp based on visual input [1], [17]. To address the risks and costs associated with real-world training, some studies have proposed training the RL module entirely on simulated data [18]. Additionally, studies incorporate domain adaptation techniques to bridge the gap between the simulation and real-world environments, aiming to achieve better performance in real-world applications [19], [20].

In 6-DoF closed-loop grasp, work with closed-loop system with multi-view design has been proposed [21]. For reinforcement learning in 6-DoF task, challenging training design due to high-dimensional action space need to be resolved. Song et al. [12] tackles this issue by action-view based rendering to improve learning efficiency. Wong et al. [9] combines RL with Imitation Learning, Behavior Cloning is incorporated into loss function to increase the similarity between the model's outputs and the demonstration data in Q-learning. [10] utilizes an encoder to compress the complete trajectory into a latent space and re-estimate grasp trajectory with closed-loop control. It utilizes an option classifier for switching to an instance grasp policy to improve performance. To reduce the required computational resources, required functions for HAGrasp are integrated into a single model.

C. Hybrid-Action RL for Robot Manipulation

RL-based methods have shown exceptional performance in complex robot tasks. However, in practical applications, many robot tasks require hybrid output control, which combines both continuous and discrete actions. Traditional RL algorithms often focus on continuous control tasks [22], [23] or discrete decision-making tasks [24] separately. This limitation poses a challenge when dealing with tasks that involve a combination of both types of actions. Advanced model design has been developed to get an optimal performance in hybrid-action RL for robotics. Zhou et al. proposed. Some works treats hybrid problems in native form, to perform a point-wise selection for contact-rich manipulation [25] or design further optimization for specific task [3].

II. Proposed Method

We formulate the problem as a Parameterized Action Markov Decision Process (PAMDP) [16] with a point cloud input and sparse reward. Our approach consists of two

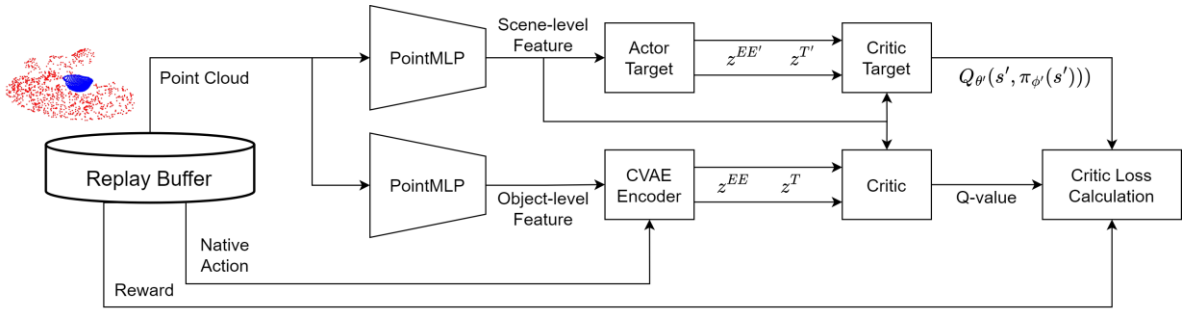


Fig. 3. The training design of deep reinforcement learning for HAGrasp policy. The critic estimates Q-value with specific state and latent action projected by CVAE encoder.

component, action-mapping CVAE and hybrid-action grasp module. Our control policy learns to predict latent actions for unseen scene in 6-DoF grasp task.

Three steps are performed for training HAGrasp:

- *Offline CVAE training*: train CVAE for action remapping with demonstration data.
- *Fine-tuning with Object-level Exploration:* improve CVAE on online dataset for higher action diversity.
- *Scene-level HAGrasp training*: update control policy for collision-aware grasp system.

A. Hybrid Action Grasp Module

To maximize the benefit from closed-loop control, the grasp module takes accumulated point clouds [9] to allow multi-view state input. The proposed system has two control action in robotic grasp tasks: end-effector control A^{EE} and termination evaluation A^T. Specifically, the gripper control action is a continuous action of 6 values (x, y, z, R_x, R_y, R_z) including relative position (x, y, z) and rotation in form of Euler angles (R_x, R_y, R_z). The termination evaluation is a discrete action of binary value $A^T \in \{0, 1\}$.

To achieve hybrid-action control, the proposed model is designed to implement Q-learning in latent action space. We adopt actor-critic architecture from TD3 [23] as our reinforcement learning basis. Fig. 3 shows the optimization process of the critic network Q_θ in hybrid action grasp module. The actor network π_ϕ of proposed approach outputs two latent actions (z^{EE}, z^T) which represent end-effector and gripper control in latent action space individually. When proposed system interacts to environments, the native gripper action is chosen by calculating pairwise distance with embedding table in CVAE, and the continuous pose action is reconstructed with decoder and object-level point cloud. To implement Q-learning in uniform action space, the inputs of the critic are projected from native action to latent action by encoder. The training target for the critic is defined as (1). The model approximate Q-value with both two vectors in hybrid action.

$$L_Q = (Q_\theta(s, z^{EE}, z^T) - y)^2 \tag{1}$$

where $y = r + \gamma Q_{\theta'}\left(s', \pi_{\phi'}(s')\right)$ with definition of Bellmen equation, s' is next state and r is the reward from rollout. $Q_{\theta'}$ and $\pi_{\phi'}$ are the target networks and γ is discount factor.

To address the challenges of high-dimensional action space, we apply behavior cloning with additional loss function to enhance the training efficiency by incorporating the data from the demonstration. We employ a sampling-based path planning algorithm [26] as the demonstration planner for the expert dataset D_{expert} in the grasp task. After selecting the suitable grasp pose with the known object dataset, point cloud and relative pose would be recorded at each waypoint. The output actions of the control policy are compared with the expert actions based on spatial distance, and loss function L_{BC} is calculated to minimize the difference between the prediction and the expert actions as (1).

$$L_{BC} = L_{pose}\left(A^{EE}_{pred}, A^{EE}_{exp}\right) + \beta_{BC} topk\left(\left\|z^T_{pred} - z^T_{exp}\right\|_2^2\right) \tag{2}$$

where A^{EE}_{pred} is reconstructed from z^{EE}_{pred} with CVAE decoder to Cartesian space as expert action A^{EE}_{exp}. To apply behavior cloning on learnable termination, average distance in latent action space between projected expert action z^T_{exp} and predicted termination z^T_{pred} is calculated. Additionally, we also take data which $A^T = 1$ from online rollout to apply this optimization as self-supervised manner. As the low portion of waypoints that fulfill termination condition would lead to performance drop due to data imbalance, only top-k largest error of termination action would be back-propagated. L_{pose} in (3) is used to define point group distance after projecting 6-DoF pose to gripper control point as defined in [27].

$$L_{pose} = \frac{1}{|X_g|}\sum_{x \in X_g} \|A^{EE}_1 - A^{EE}_2\|_1 \tag{3}$$

After dataset from online rollout is collected, training data are sampled from both D_{expert} and D_{online}. The policy gradient is use to maximize expect return with output from the critic. The final loss function of HAGrasp is combination of policy gradient and behavior cloning as in (4). The design of behavior cloning improves the efficiency of training process and perform better collision-awareness than only optimized by policy gradient.

$$L_{policy} = -Q_\theta\left(s, \pi_\phi(s)\right) + L_{BC} \tag{4}$$

The process of offline collection and online rollout are executed in simulated environments, which brings risk of performance drop due to sim-to-real gap due to deformation of input point clouds across different domains [28]. To address the sim-to-real challenges, we adopt several design choices to improve training process. Firstly, the model input is chosen to be point clouds instead of RGB-D images. Point clouds demonstrate better generalization for 6-DoF grasp with geometric feature in the test environment compared to RGB-D input [9], [29]. Secondly, data augmentation is employed to preprocess the training data. The original point clouds are subjected to random perturbations with jittering, to introduce noise and simulate real-world data.

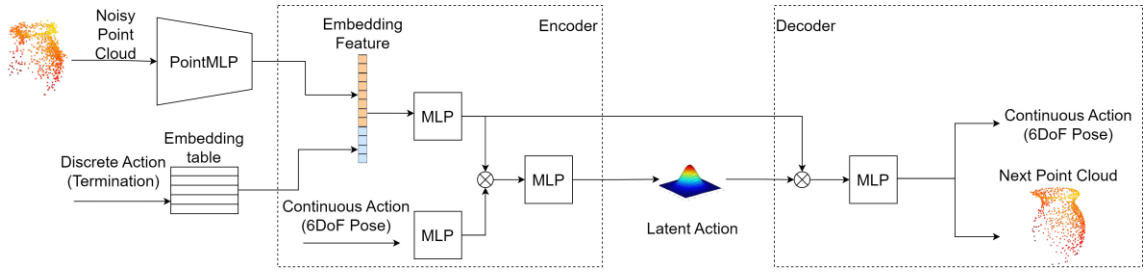


Fig. 4. Network design of action-mapping conditional variational autoencoder.

B. Action-Mapping using Conditional VAE

To implement the hybrid-action control of HAGrasp, an encoder is used to map the latent value for training, while a decoder is employed to reconstruct the actions for actual control. As shown in Fig. 4, we implement our action-mapping CVAE for latent action spaces, continuous actions are used as inputs and reconstruction objectives for the CVAE. Discrete actions and point cloud features are concatenate into the model training as conditions. The discrete actions are treated as trainable parameters and translated into feature vectors using an embedding table, which indexes original actions and converts them into latent one. To maintain stable performance for CVAE, object-level point cloud features that extracted by PointMLP [30] is used as conditional feature. The decoder is designed to predict the point cloud of next step as an additional task. It is used to enhance the understanding of the interaction between actions and the environment and have stable encoding for RL training. Further, feature extractor can learn representation of denoising by point cloud reconstruction from noisy input. The Chamfer distance L_{CD} is used as loss function for deformation reconstruction [28]. The CVAE predicts the denoised version of next point cloud S_{pred} based on the decoded action and optimize with output the ground truth S_{t+1}.

The expert data D_{expert} are collected for both Behavior Cloning and CVAE training. By utilizing pre-defined grasp poses and trajectory planning with a high success rate, this initialization makes CVAE output reasonable actions during the training phase of policy learning and reduces the required time for the control policy to explore. The final loss function (5) for CVAE is consist of weighted KL-divergence regularization loss [31] $L_{KL} = KL(\mathcal{N}(\mu(A,z),\sigma((A,z))) \parallel \mathcal{N}(0,I))$ and reconstruction loss.

$$L_{CVAE} = \|A_1^{EE} - A_2^{EE}\|_2^2 + L_{pose}(A_1^{EE}, A_2^{EE}) + \beta_{KL} L_{KL} + L_{CD}(s_{pred}, s_{t+1}) \quad (5)$$

the weight of KL loss β_{KL} is dynamical adjusted to stabilize KL-divergence at pre-defined value [32].

C. Fine-tuning with Object-level Exploration

Offline training of CVAE enables the encoder and decoder to have a certain level of generalization. There can still be limitations in terms of diversity in the continuous data space, even with a large amount of expert-collected training data. This limitation arises due to the imbalance and lack of diversity in the spatial distribution of the expert dataset, which can result in a poor generalization in the trained model.

To improve generalization of proposed approach, the training design involves joint training with object-level grasp task. The CVAE and the single-object grasp module are simultaneously optimized. By optimizing the control policy of object-level grasp module to improve the success rate, and at the same time, the diversity of the CVAE dataset is increased. As shown in Fig. 5, the grasp module parameters are updated during exploration in the collection process of online dataset D_{online}. Additionally, due to the absence of occlusions on the target object, the model is able to accurately locate the grasp pose that is closest to the end effector. In order to enhance the robustness of the model, the design includes the prediction of the final grasp pose G_f as [9] with $L_{goal} = L_{pose}(G_{pred}^f, G_{gt}^f)$.

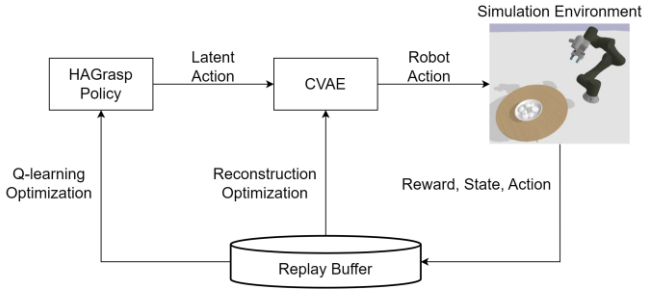


Fig. 5. The object-level grasping module collects online data with trained CVAE through interaction with simulated environment, then save in replay buffer. The data would be sampled from replay buffer as training data for both CVAE and object-level HAGrasp.

For reward in grasp task, to enhance the geometry-aware representation of termination evaluation learned by policy, penalty are assigned to actions that lead to failed grasp as $-L_{pose}(G_{EE}, G_f)$. These penalty values are determined with the distance between the current end-effector pose G_{EE} and nearest grasp pose G_f. A positive reward is given at the step of successful grasp. To collect higher-quality data in low-difficulty tasks, the penalty value for end-effector collision is set to its maximum. These designs are made with the intention that the control policy generates actions that do not result in collisions with objects while ensuring a high success rate.

D. Scene-level HAGrasp

To optimize for more complex object placements and environment, the input data of the HAGrasp includes point clouds of both obstacles and target objects, allowing the model to have a more comprehensive understanding of the scene. Additionally, instance segmentation is used to obtain object masks, which indicates the target object in a binary form.

Due to the different property of the tasks, the design of the scene-level grasp module differs in several aspects. Firstly, the complexity in space increases significantly, making it challenging to infer the target pose at each waypoint solely based on a predefined final grasp pose. Therefore, the design of predicting the final grasp pose is not used in this case. Second, through our observations in experiment, it was found that setting excessively high penalty values for collisions would cause the control policy to underestimate the expected

TABLE I
EVALUATION STATISTICS IN SIMULATIONS OF 7 STABLE OBJECTS, CD DENOTES COLLISION DETECTION

Method	7 objects		
	Success rate	Inference time	Cycle time
6-DoF Graspnet [4] + CD	59.6%	3.15s	8.4s
Contact-Graspnet [5]	71.4%	0.6s	5.2s
HAGrasp-eva	76.2%	0.07s	5.6s
HAGrasp (proposed)	**81.1%**	**0.055s**	**5.1s**

reward in a cluttered environment. This could result in the policy prioritizing collision avoidance over grasping. As definition in (6), the penalty values for collisions are reduced and dynamically adjusted based on the distance to the target.

$$r = \begin{cases} 1, & if\ successful\ grasp \\ -L_{pose}(G_{EE}, G_f), & if\ failed\ grasp \\ -L_{pose}(G_{EE}, G_f), & if\ collided \\ 0, & otherwise \end{cases} \quad (6)$$

This adjustment aims to encourage the control policy to consider the potential risk of slight contact and prioritize reaching deep into the graspable region when selecting the strategy that maximizes the expected reward. By considering these task-specific design adjustments, the scene-level grasp module can effectively adapt to the increased complexity of the environment and make informed decisions that balance obstacle avoidance and grasp success.

III. SIMULATION AND EXPERIMENTAL RESULTS

We evaluate HAGrasp on a benchmark against 2 open-loop grasp baseline methods, 6-DoF Graspnet [4] and Contact-Graspnet [5]. A version without hybrid action termed HAGrasp-eva is first designed as a temporally model for comparative study. HAGrasp-eva is used to indicate a waypoint-only version of HAGrasp which is trained by script termination condition and uses evaluator from [4] in test time. HAGrasp uses learnable termination design with hybrid action design. TM5-900M, a 6-DoF robotic arm with a Realsense D435 mounted on robot gripper is used to perform the grasp task. For each experiment, random category of objects in different number would be chosen and set random pose in workspace. After each episode of grasping in one experiment, the target object would be removed either success or not.

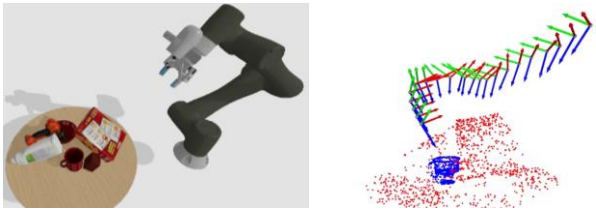
Both CVAE and RL models are trained with simulated data from Acronym [33]. During scene-level training, 3 to 7 objects from 1333 categories were randomly stacked in simulation scene for grasp task to ensure generalization to wider range of practical applications. The feature exactor takes point cloud as input and output 1024 dimension features. The CVAE encoder compacts actions to 64 dimension latent value, and the value of predefined KL-divergence is set to 15. The discount factor γ is set to 0.97. β_{BC} is set to 5.

A. Simulation Results of Unseen Object Grasp

The proposed method is evaluated in Pybullet [34] with 20 different YCB objects [35] and execute 100 round experiments. We design two environment setup to address variation on complexity, the *stable* arrangement placed object with random pose where has no collisions with other objects. In the *cluttered* arrangement, objects are randomly dropped into the scene, and their poses are fixed after colliding with the scene or other objects. Fig. 6(a) illustrates objects in *cluttered* setup,

TABLE II
PERFORMANCE IN SIMULATION ENVIRONMENT OF CLUTTERED OBJECTS, CD DENOTES COLLISION DETECTION

Method	Cluttered		
	3 objects	5 objects	7 objects
	Success Rate		
6-DoF Graspnet [4] + CD	58.3%	43.8%	36.2%
Contact-Graspnet [5]	74.0%	68.5%	66.4%
HAGrasp-eva	80.0%	74.4%	71.6%
HAGrasp (proposed)	**82.1%**	**79.7%**	**77.6%**



(a) (b)

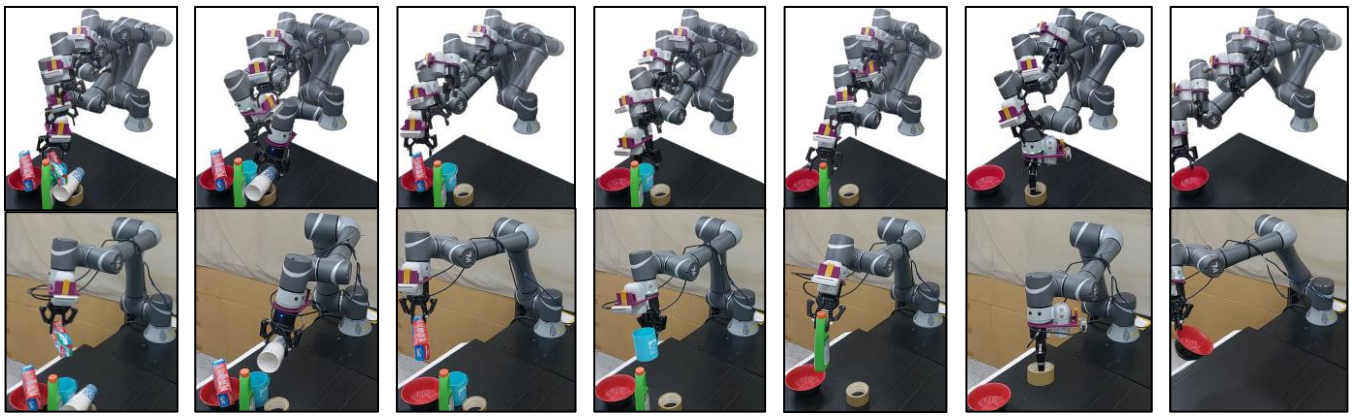
Fig. 6. Experiment result of simulation in *cluttered* arrangement. (a) object placement and initial pose of robotic arm, (b) the complete trajectory of grasp task.

which are randomly stacked in simulation. As shown in Fig. 6(b), with integrating dynamic adjustment and hybrid output into a single model, our approach successfully plans trajectories and achieves grasp success in collision-free spaces.

Table I shows the simulation results in *stable* arrangement. This simulation aims to verify the effectiveness of closed-loop system and performance improvement brought by proposed hybrid-action RL design. Open-loop methods 6-DoF Graspnet suffers from performance drop due to their inability to avoid obstacles between the initial and target points. On the other hand, Contact-Graspnet may not consider collision avoidance for each trajectory waypoint, leading to potential collisions while moving to pre-grasp positions.

For the simulation of *cluttered* arrangement, Table II shows that the improvement in grasp success rate becomes more evident as the complexity of the environment increases. The main factors contributing to the performance drop to open-loop methods were occlusions caused by random object poses. Relying on single-view input for estimation may result in estimation errors due to occluded obstacles or hidden grasps caused by fragmented point clouds. The proposed system overcomes this challenge by incorporating multi-view inputs, continuously accumulating environmental information during runtime. It improves the overall system performance by 77.6% success rate while ensuring efficient execution and reduced grasp task cycle time. HAGrasp is capable of executing control strategies with higher expected success rate based on current task requirements, achieving a balance among strategies for obstacle avoidance, grasping, and environmental information acquisition, achieving stable performance in cluttered scene.

The HAGrasp-eva shows advantages of multi-view input and closed-loop control. Results also show that it lacks of understanding of multi-modality of grasp pose which verifies the effectiveness of hybrid action in reward design. Due to external evaluator which takes different representation, resulting in numerous unnecessary collisions in refining grasp poses. The full design of HAGrasp achieves 81.1% success rate on 7 objects simulation while closed-loop baseline with external model only has 76.2%. The improvement in success rate is attributed to the shared features among the hybrid action of HAGrasp which enables a comprehensive understanding.



(a) (b) (c) (d) (e) (f) (g)

Fig. 7. Experimental results in real world for picking 7 novel objects. (a) trajectory and the robot pose of grasp success in first episode, (b)(c)(d)(e)(f)(g) shows the process of completing the task and grasping the other 6 objects.

B. Real Robot Experiments

In real robot experiments, ROS is used for communication between the robot and the sensors. The target mask for novel objects is obtained using the UOAIS [36]. To validate the proposed architecture, the grasp system was tested in real-world scenarios with noisy data from real sensor and random object poses in a cluttered environments. 20 grasp experiments were conducted with 20 daily life unseen objects.

Fig. 7(a) shows the experimental result of grasping the first target of 7 unseen objects, the upper image is the trajectory of complete one episode of grasp, and below is the robot pose that successfully grasp. The process of successfully grasping the other 6 objects in sequence are shown in Fig. 7(b) to Fig. 7(g). As shown in Fig7(b), the grasp pose is determined as the bottom part of the bottle avoiding the potential collision with near by objects.

The real world experiment of 5 and 7 objects are conducted to verify efficiency and robustness of proposed hybrid action control system. As shown in Table III, we compare the overall performance of HAGrasp with open-loop baseline Contact-Graspnet [5] which is a scene-level approach that estimate collision-free grasp pose in real world. The experimental results of grasp 5 objects in clutter show that the success rate of the proposed HAGrasp is 78.0%, while baseline method is 64.0%. Our method has a improvement of 21.8% in success rate compared to the open-loop control method.

For the grasp task of 7 unseen objects, the HAGrasp outperforms Contact-Graspnet by 20.8% improvement on success rate. It is observed that Contact-Graspnet has poor performance on grasping small objects due to its one-stage architecture. On the other hand, our approach takes scene-level to perform collision-free motion planning and further implements target-centric grasp with object-level feature. The open-loop baseline suffers more from sensor noise cause by reflection and occlusion, the HAGrasp achieves more stable performance with dynamical adjustment using closed-loop control. The proposed method executes data preprocessing and re-estimation at every step which reduce the effect of sensor noise. Regarding the effects of multi-view input and sim-to-real design, the HAGrasp can take noisy point cloud as input and perform success grasp. The performance achieved in real-world experiments, which closely matches the results obtained in simulated experiments. It demonstrates the effectiveness of our sim-to-real training design and validates the model's ability to generalize effectively across different domains and avoid performance drop when applied to real-world data.

TABLE III
PERFORMANCE OF REAL ROBOT EXPERIMENTS

Method	5 obj.	7 obj.		
	Success rate		Cycle time	Inference time
Contact-Graspnet [5]	64.0%	61.4%	29s	0.6s
HAGrasp (proposed)	**78.0%**	**74.2%**	**27s**	**0.055s**
Improvement	**21.8%**	**20.8%**	**6.8%**	**90.8%**

Our approach has great improvement in inference time, the average computation time of one estimation is *0.055s* where open-loop baseline has *0.6s* inference time. The HAGrasp simplifies grasp task to waypoint predict instead of grasp estimation of whole scene, which has faster inference with fully capability of closed-loop control. The cycle time of propose method is shorter than baseline, which proves model efficiency of our design. The cycle time of closed-loop control is decreased due to multiple estimations for object mask, it can be further improved with optimal segmentation design.

IV. CONCLUSIONS AND FUTURE WORK

We present a closed-loop grasp control system based on deep reinforcement learning to perform hybrid action robot grasp. A single-model design to integrate both the functions of end-effector control and termination evaluation has been proposed, which improves the efficiency of the autonomous grasp system. Furthermore, the closed-loop control with hybrid-action design is employed to further enhance the grasp success rate. Our work outperforms baseline approach by 90.8% improvement on inference time with hybrid-action grasp network. Results shows that the proposed HAGrasp improve grasp success rate by 21.8% in 5 objects cluttered scene while reduce computational load with 6.8% cycle time improvement. Experimental results show that HAGrasp learns from sim-to-real training design by achieving stable performance of 74.2% success rate for grasping 7 objects in real world. The future work will focus on development of image segmentation process to further improve cycle time.

ACKNOWLEDGMENT

We would like to thank Techman Robot Inc. for the support of this project and Mr. Shao-Heng Chien and Mr. Wei-Han Chen for their assistance in the robot experiments.

REFERENCES

[1] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke and S. Levine, "QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation," in *Proc. 2nd Conference on Robot Learning*, Zürich, Switzerland, 2018, pp. 651-673.

[2] M. Breyer, F. Furrer, T. Novkovic, R. Siegwart and J. Nieto, "Comparing Task Simplifications to Learn Closed-Loop Object Picking Using Deep Reinforcement Learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, 2019, pp. 1549-1556.

[3] M. Neunert, A. Abdolmaleki, M. Wulfmeier, T. Lampe, J. T. Springenberg, R. Hafner, F. Romano, J. Buchli, N. Heess and M. A. Riedmiller, "Continuous-discrete Reinforcement Learning for Hybrid Control in Robotics," in *Proc. 3rd Annual Conference on Robot Learning*, Osaka, Japan, 2019, pp. 735-751.

[4] A. Mousavian, C. Eppner and D. Fox, "6-DOF GraspNet: Variational Grasp Generation for Object Manipulation," in *Proc. 2019 IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, 2019, pp. 2901-2910.

[5] M. Sundermeyer, A. Mousavian, R. Triebel and D. Fox "Contact-graspnet: Efficient 6-DoF Grasp Generation in Cluttered Scenes," in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, X'ian China, 2021, pp. 13438-13444.

[6] C. Wu, J. Chen, Q. Cao, J. Zhang, Y. Tai, L. Sun and K. Jia, "Grasp Proposal Networks: An End-to-End Solution for Visual Learning of Robotic Grasps," in *Proc. 34st Conference on Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 13174-13184.

[7] C. Wang, H. -S. Fang, M. Gou, H. Fang, J. Gao and C. Lu, "Graspness Discovery in Clutters for Fast and Accurate Grasp Detection," in *Proc. 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, 2021, pp. 15944-15953.

[8] H. Ma and D. Huang, "Towards Scale Balanced 6-DoF Grasp Detection in Cluttered Scenes," in *Proc. 6th Conference on Robot Learning*, Auckland, New Zealand, 2022, pp. 2004-2013.

[9] L. Wang, Y. Xiang and D. Fox, "Goal-Auxiliary Actor-Critic for 6D Robotic Grasping with Point Clouds," in *Proc. Conference on Robot Learning*, London, UK, 2021, pp. 70-80.

[10] L. Wang, X. Meng, Y. Xiang and D. Fox, "Hierarchical Policies for Cluttered-Scene Grasping with Latent Plans," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2883-2890, 2022.

[11] D. Yang, T. Tosun, B. Eisner, V. Isler and D. Lee, "Robotic Grasping through Combined Image-Based Grasp Proposal and 3D Reconstruction," in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, X'ian China, 2021, pp. 6350-6356.

[12] S. Song, A. Zeng, J. Lee and T. Funkhouser, "Grasping in the Wild: Learning 6DoF Closed-Loop Grasping from Low-Cost Demonstrations," *Robotics and Automation Letters*, vol. 5, no. 3, pp. 4978-4985, 2020.

[13] A. Allshire, R. Martín-Martín, C. Lin, S. Manuel, S. Savarese and A. Garg, "LASER: Learning a Latent Action Space for Efficient Reinforcement Learning," in *Proc. IEEE International Conference on Robotics and Automation*, Xi'an, China, 2021, pp. 6650-6656.

[14] B. Li, H. Tang, Y. Zheng, J. Hao, P. Li, Z. Wang, Z. Meng and L. Wang, "HyAR: Addressing Discrete-Continuous Action Reinforcement Learning via Hybrid Action Representation," in *Proc. 10th International Conference on Learning Representations*, Virtual Event, 2022, pp. 1-22.

[15] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. 2nd International Conference on Learning Representations*, Banff, AB, Canada, 2014, pp. 1-14.

[16] W. Masson, P. Ranchod and G. D. Konidaris, "Reinforcement Learning with Parameterized Actions," in *Proc. Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, USA, 2016, pp. 1934-1940.

[17] E. Jang, C. Devin, V. Vanhoucke and S. Levine, "Grasp2Vec: Learning Object Representations from Self-Supervised Grasping," in *Proc. 2nd Conference on Robot Learning*, Zürich, Switzerland, 2018, pp. 99-112.

[18] S. Iqbal, J. Tremblay, A. Campbell, K. Leung, T. To, J. Cheng, E. Leitch, D. McKay and S. Birchfield, "Toward Sim-to-Real Directional Semantic Grasping," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 2020, pp. 7247-7253.

[19] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell and K. Bousmalis, "Sim-to-Real via Sim-to-Sim: Data-Efficient Robotic Grasping via Randomized-to-Canonical Adaptation Networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019, pp. 12627-12637.

[20] K. Rao, C. Harris, A. Irpan, S. Levine, J. Ibarz and M. Khansari, "RL-CycleGAN: Reinforcement Learning Aware Simulation-to-Real," in *Proc. the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 11154-11163.

[21] D. Son, "Grasping as Inference: Reactive Grasping in Heavily Cluttered Environment," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7193-7200, 2022.

[22] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver and D. Wierstra, "Continuous Control with Deep Reinforcement Learning," in *Proc. 4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016, pp. 1-14.

[23] Scott Fujimoto, Herke van Hoof and David Meger, "Addressing Function Approximation Error in Actor-Critic Methods," in *Proc. the 35th International Conference on Machine Learning*, Stockholm, Sweden, 2018, pp. 1582-1591.

[24] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *arXiv preprint arXiv:1312.5602*, 2013.

[25] W. Zhou, B. Jiang, F. Yang, C. Paxton and D. Held, "HACMan: Learning Hybrid Actor-Critic Maps for 6D Non-Prehensile Manipulation," in *Proc. 7th Conference on Robot Learning*, Atlanta, Georgia, USA, 2023, pp. 241-265.

[26] J. J. Kuffner and S. M. LaValle, "RRT-Connect: An Efficient Approach to Single-Query Path Planning," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, San Francisco, CA, USA, 2000, pp. 995-1001.

[27] Y. Li, G. Wang, X. Ji, Y. Xiang and D. Fox, "DeepIM: Deep Iterative Matching for 6D Pose Estimation," in *Proc. the European Conference on Computer Vision*, Munich, Germany, 2018, pp. 695-711.

[28] I. Achituve, H. Maron and G. Chechik, "Self-Supervised Learning for Domain Adaptation on Point Clouds," in *Proc. the IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, 2021, pp. 123-133.

[29] X. Yan, M. Khansari, J. Hsu, Y. Gong, Y. Bai, S. Pirk and H. Lee, "Data-Efficient Learning for Sim-to-Real Robotic Grasping using Deep Point Cloud Prediction Networks," *arXiv preprint arXiv:1906.08989*, 2019.

[30] X. Ma, C. Qin, H. You, H.i Ran and Y. Fu, "Rethinking Network Design and Local Geometry in Point Cloud: A Simple Residual MLP Framework," in *Proc. The Tenth International Conference on Learning Representations*, Virtual Event, 2022, pp. 1-15.

[31] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed and A. Lerchner, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework," in *Proc. 5th International Conference on Learning Representations*, Toulon, France, 2017, pp. 1-22.

[32] H. Shao, S. Yao, D. Sun, A. Zhang, S. Liu, D. Liu, J. Wang and T. F. Abdelzaher, "ControlVAE: Controllable Variational Autoencoder," in *Proc. the 37th International Conference on Machine Learning*, Virtual Event, pp. 8655-8664.

[33] C. Eppner, A. Mousavian and D. Fox, "Acronym: A Large-Scale Grasp Dataset Based on Simulation," in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, X'ian, China, 2021, pp. 6222-6227.

[34] E. Coumans and Y. Bai, Pybullet, a Python Module for Physics Simulation for Games, Robotics and Machine Learning, 2016.

[35] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. Dollar, "Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols," *IEEE Robotics and Automation Magazine*, vol. 22, no. 3, pp. 36-52, 2015.

[36] S. Back, J. Lee, T. Kim, S. Noh, R. Kang, S. Bak and K. Lee, "Unseen Object Amodal Instance Segmentation via Hierarchical Occlusion Modeling," in *Proc. 2022 International Conference on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, 2022, pp. 5085-5092.